



LIBRARY
OF THE
UNIVERSITY
OF ILLINOIS

370

Il6t

no. 1-5

~~EDUCATION~~

Return this book on or before the
Latest Date stamped below.

University of Illinois Library

DEC 31 1970

DEC 13 1976


DEC 22 1976

MAY 25 1982

DEC 22 1983

MAY 19 1984

L161—H41



Digitized by the Internet Archive
in 2012 with funding from
University of Illinois Urbana-Champaign

<http://www.archive.org/details/someempiricalstu04rudi>



BUREAU OF RESEARCH AND SERVICE
College of Education
University of Illinois
Urbana, Illinois

**THE RELATIONSHIP OF INTERPERSONAL PERCEPTION TO
EFFECTIVENESS IN BASKETBALL TEAMS**

Fred E. Fiedler, Walter Hartmann, and Stanley A. Rudin
College of Education, University of Illinois

Technical Report No. 3

- - - -

**SOME EMPIRICAL STUDIES OF THE RELIABILITY OF
SOCIAL PERCEPTION SCORES**


Stanley A. Rudin, Irving Lazar, Mary E. Ehart and Lee J. Cronbach
College of Education, University of Illinois

Technical Report No. 4

**Studies performed under Contract N6ori-07135
with the Office of Naval Research**

**Project on
Social Perception and Group Effectiveness**

May, 1952

A large redacted area in the top right corner of the page, consisting of a solid black rectangle.

SOME EMPIRICAL STUDIES OF THE RELIABILITY OF
INTERPERSONAL PERCEPTION SCORES

Stanley A. Rudin, Irving Lazar, Mary E. Ehart, and Lee J. Cronbach
College of Education, University of Illinois

Study performed under Contract N6ori-07135
with the Office of Naval Research

Project on
Social Perception and Group Effectiveness

Technical Report No. 4
May, 1952

SOME EMPIRICAL STUDIES OF THE RELIABILITY OF INTERPERSONAL PERCEPTION SCORES * ^{1/}

Stanley A. Rudin, Irving Lazar, Mary E. Ehart, and Lee J. Cronbach
College of Education, University of Illinois

Introduction

Fiedler and others have suggested that Assumed Similarity, Real Similarity, and other measures of social perception represent psychological attributes influencing group behavior. Each of these variables is operationally defined in terms of the similarity between two descriptions of individuals on a set of items. To demonstrate that these measures of social perception are meaningful and general concepts, they must be shown to have reliability or consistency.

This study is an attempt to determine whether such reliable individual differences exist. Specifically, it is an investigation of the reliabilities of the measures used in a study of effectiveness of basketball teams, reported by Fiedler, Hartmann, and Rudin (3). Consistency of the similarity measures can be demonstrated in these ways:

- (1) over items;
- (2) over persons ("perceivees"; the objects of perceptions);
- (3) over time.

The first two methods have been applied to Assumed Similarity scores in this study. Only the first type of reliability was studied for Real Similarity

*This study was made in connection with Contract N6ori-07135 between the University of Illinois and the Office of Naval Research, Human Relations Branch. Technical Report No. 4, May 1952.

^{1/} The authors wish to acknowledge with pleasure the cooperation of W. G. Warrington in solving many computational problems.

scores. Stability over time has not been investigated.

Development of the Instrument

Since the test was constructed under time pressure there was not sufficient time to run a pilot study, to pretest items, or thoroughly to investigate the assumptions involved in the various types of instrument that might have been used.

On the basis of a previous study by Fiedler, it was decided to use a modification of the Stephenson Q-technique(5). Briefly, the instrument was devised in the following manner:

- (1) 100 statements descriptive of a person's behavior were derived from interviews with high school students. The statements were then grouped into 20 blocks of 5 items each.^{2/} The items dealt with a variety of personality traits.
- (2) The subject was instructed to indicate which statement in each block he considered to be most characteristic of himself, and which statement in each block he considered least characteristic of himself.

The subjects were high school basketball players from 14 teams, totaling 178 subjects. For a more detailed description of the subjects and procedures, see the paper referred to previously (3).

Measures

The instrument was administered to each subject four times, with these instructions:

- (1) Mark the items to describe yourself (self - s)
- (2) Mark the items to describe how you would ideally like to be (ideal - i)
- (3) Predict how the person with whom you can cooperate best will describe himself ("positive" choice - p)
- (4) Predict how the person with whom you can cooperate least well will describe himself ("negative" choice - n).

^{2/} More recent work in this project suggests that the five-item block is less efficient than Stephenson's sorting technique. Studies of this problem are still incomplete.

Each series of responses could be correlated with another to obtain an index of the similarity with which the two tests have been marked. This was done by comparing each block in one test with the same block in the other test and assigning a numerical score to each pair of blocks. Five degrees of similarity were possible. Perfect agreement was counted as +2, an agreement of only one pair of the markings as +1, neither agreement or disagreement as 0, a disagreement of only one pair of the markings as -1, and disagreement of both pairs of markings (perfect disagreement) as -2. Examples:

Perfect Agreement		Partial Agreement		Perfect Disagreement	
Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
Most	Least	Most	Least	Most	Least
1	X	X		X	
2					
3			X		
4					
5	X	X		X	X
Scored +2		Scored +1		Scored -2	

The scores obtained in this way are a linear function of the Pearson product-moment correlation, when block scores are added and divided by 40.

It is obvious that two classes of correlations can be obtained: (1) correlations between two tests taken by the same subject but with different instructions (the "internals"), and (2) correlations between two tests each taken by two different subjects ("externals"). The internals or intrapersonal correlations yield measures of Assumed Similarity (AS); i.e., a measure of how similar the subject perceives himself to be to someone else, or how similar he perceives two other persons to be. The externals or interpersonal correlations yield a measure of "Real Similarity (RS), which is a measurement of two persons' actual similarity on these items.

The reliability of the following measures was determined in the present study:

Internals

- (1) AS_p: the assumed similarity between self and the prediction of the best co-worker (s vs. p)
- (2) AS_n: the assumed similarity between self and the prediction of the person with whom the subject said he could cooperate least well (s vs. n)

Each group of students was assigned a specific area of the map to study. The students were to identify the major features of the area, including the location of the major cities, the major rivers, and the major mountains. The students were to prepare a report on their findings, and to present their report to the class. The students were to be prepared to answer questions about their report, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class.

Group 1		Group 2		Group 3	
Area	Features	Area	Features	Area	Features
1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36
37	38	39	40	41	42
43	44	45	46	47	48
49	50	51	52	53	54
55	56	57	58	59	60
61	62	63	64	65	66
67	68	69	70	71	72
73	74	75	76	77	78
79	80	81	82	83	84
85	86	87	88	89	90
91	92	93	94	95	96
97	98	99	100	101	102

The students were to be prepared to answer questions about their report, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class. The students were to be prepared to defend their findings, and to discuss their findings with the class.

The students were to be prepared to answer questions about their report, and to discuss their findings with the class.

The students were to be prepared to answer questions about their report, and to discuss their findings with the class.

The students were to be prepared to answer questions about their report, and to discuss their findings with the class.

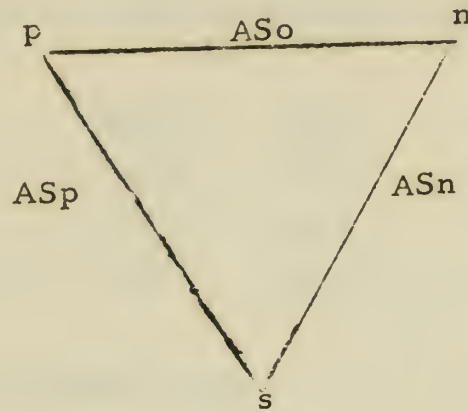
The students were to be prepared to answer questions about their report, and to discuss their findings with the class.

The students were to be prepared to answer questions about their report, and to discuss their findings with the class.

The students were to be prepared to answer questions about their report, and to discuss their findings with the class.

- (3) ASo: the assumed similarity between the person's best co-worker and the person with whom he said he could cooperate least well (p vs. n).

These measures are schematized in Figure 1.



s = self description

p = prediction of best co-worker

n = prediction of the person with whom S can
cooperate least well

SCHEMATIC DIAGRAM OF THE MEASURES USED IN THIS STUDY

Figure 1

Externals

RS: the "real similarity" between the self descriptions of two persons on this test.

Internals.

Method. Seventy-seven players were chosen at random from the total N of 178. The test was split into odd-even halves of 10 blocks each. The internal similarity score over odd blocks was compared with the similarity score over even blocks by means of Guttman's split-half formula (4).

Results. Correlations between the three internal measures are shown in Table 1.

Table 1

CORRELATIONS BETWEEN THE INTERNAL MEASURES (N = 70)

<u>Variables</u>	<u>r</u>
ASp vs. ASo	.18
ASp vs. ASn	.30
ASn vs. ASo	.62

A summary of the reliabilities can be found in Table 2. Inspection of distributions of the three measures showed them to be fairly normal. All these correlations are significantly greater than zero ($P < .01$).

Table 2
RELIABILITIES OF THE MEASURES OF ASSUMED SIMILARITY
(N = 77)

<u>Variable</u>	<u>Mean*</u>	<u>SD*</u>	<u>Reliability</u>
ASp	.297	.179	.618
ASn	.097	.200	.695
ASo	.111	.214	.608

*In terms of the Q correlation scale

Conclusions. The basketball instrument is sufficiently reliable to detect group differences in AS and reliable enough for investigations similar to the basketball study. On the other hand, greater reliability would be necessary before the instrument could be used for drawing conclusions about single individuals.

The fact that AS scores are consistent from one set of items to another supports the hypothesis that this aspect of social perception is a trait in which consistent individual differences exist. The establishment of AS as a measurable trait justifies further investigation to determine its psychological meaning and practical significance.

Externals.

Method. While it was considered desirable to find the reliability of interpersonal measures, correlating the similarity scores of odd vs even blocks for each pair among 77 persons (2,926 comparisons) would have involved too much time and effort. Furthermore, mathematical theory indicates (cf. Technical Report No. 2) that the RS measures for different persons would have different reliability, depending on the variability of the traits measured within the person. Therefore, the reliability of RS scores was determined separately for each of five key subjects selected from one school. The five key subjects were selected at random.

Their self descriptions were split into odd-even halves, by blocks. These halves were then compared with the corresponding halves of each of the 77 subjects in the above-mentioned sample, and similarity scores summed for each half in the same manner as for the internal measures. Thus, each reliability

coefficient indicates the consistency of the instrument in reporting which of the 77 subjects are most similar to the one key person.

Results. Table 3 presents the reliability of real similarity scores involving the five key persons used. As can be seen, all are quite low, and three of the five are negative. The negative reliabilities may, for the purposes of the present study, be regarded as chance departures from zero.

Table 3
RELIABILITIES OF REAL SIMILARITY TO EACH OF FIVE KEY
PERSONS (N = 77)

<u>Person</u>	<u>Reliability</u>
1	-.226
2	.068
3	.264
4	-.571
5	-.302

Conclusions. On the present instrument, measures of interpersonal similarity are almost entirely lacking in reliability. With such scores, there is no possibility of establishing any correspondence between RS and any measure of group effectiveness.

Attempts to Improve Reliability

Two methods were used in an attempt to improve reliability: (1) eliminating items which did not correlate with other items, and (2) scoring the test in terms of clusters of items before computing similarity scores.

Items were therefore grouped into clusters. Three judges working together classified the items into 13 trial clusters on the basis of "traits" or "factors" which they were presumably tapping. Keys for each cluster were made up, and 100 persons (including the 77 used in the RS and AS studies referred to above) were scored on each of the 13 clusters. (In scoring, the item marked "most characteristic" was counted as +1, and the item marked "least characteristic as -1.) The internal consistency of each cluster was determined by means of coefficient alpha (1), modified to take into account the spurious correlation between items appearing in the same block. These alphas ranged from .48 to -.10, with a median of .21. In addition, tetrachoric r's between clusters were computed.

Clusters were then discarded if they had too low reliability and combined with others if they had high intercorrelation. After application of these criteria, eight clusters were retained and improved by item analysis. In this manner, the test was reduced to a 59 item test of eight clusters. The reliabilities (internal consistencies) of these eight clusters was estimated by using the Guttman split-half formula on a new sample consisting of the remaining 78 cases from the total N. They are shown in Table 4.

Table 4
NAMES AND RELIABILITIES OF THE EIGHT REVISED CLUSTERS
(N = 78)

<u>No.</u>	<u>Name</u>	<u>Reliability (Split half)</u>
1	Leadership	.554
2	Friendly attitude	.238
3	Comedian	.262
4	Politeness, conformity	.192
5	Competitive	.385
6	Academic, intellectual	.299
8	Aggressiveness	.596
10	Independence	.348

The clusters were originally derived in order that a profile for each person could be obtained. These profiles could then be compared, and the similarity between any two profiles indicated by the statistic D (cf. 2). Profile scoring involved the following steps: (1) each person's score on each cluster was obtained; (2) the two profiles were compared by using the formula:

$$D = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_k - b_k)^2}$$

Here a_1 = score of person a on cluster 1

b_1 = score of person b on cluster 1

a_2 = score of person a on cluster 2

b_2 = score of person b on cluster 2

D = distance (or dissimilarity) measure

The 78 new subjects who were used to find the reliabilities of the eight revised clusters were also used in this second part of the study.

Internals.

Only ASp was treated, since it was assumed that an improvement in reliability of one internal score would be indicative of the success or failure

of the profile-scoring technique in this instance.

Method. An odd-even split could not be used, since only 59 of the original 100 items were retained, thereby making some of the blocks incomplete. Therefore, the test was split into two halves with an equal number of blocks in each half. The split was arranged so that each cluster was represented by an approximately equal number of items in each half. Also, no cluster was represented by more than one item in any one block. This will be referred to as the "A-B split".

In order to test the efficiency of profile scoring, it was necessary to obtain reliabilities of ASp when scored by the profile method and when scored unclustered. Two methods were therefore used to obtain reliability: (1) the "unclustered" method, which was the same as the procedure used on the original 100-item test, except that the A-B split was used instead of an odd-even split, and (2) the "profile-scoring" method. This second method involved the following steps: each of the eight clusters was split into halves on the basis of the A-B split mentioned above. Thus, when a half A key and a half B key for each cluster were applied to a self description, it was possible to get one profile made up of all eight A halves, and another profile consisting of all eight B halves. This procedure was duplicated on the prediction of the best co-worker. This gave four profiles for each person; the A and B profiles on the self description and the A and B profiles of the prediction of the best co-worker. The A profile of the self description and the A profile of the prediction were then compared by the previously-given dissimilarity formula, and D obtained. The two B profiles were compared in the same way. This procedure yielded two assumed similarity scores, one on the A half, and one on the B half, for each person. Correlating these scores yielded the reliability coefficient.

Because of the labor involved in these computations, a sample of 40 was drawn from the group of 78 subjects and used in these calculations.

Results. A summary of the results is presented in Table 5. It can be seen that both reliabilities obtained from the shortened test are substantially higher than the reliability obtained from the original test. Significance of these differences was not tested.

Conclusions. These results suggest that elimination of ambiguous items or items uncorrelated with others in the test is beneficial, but that profile scoring of this instrument does not improve reliability. They also tend to confirm the conclusion that ASp can be reliably measured.

Table 5
RELIABILITIES OF AS_P OF THE ORIGINAL, SHORTENED,
AND PROFILE SCORED TESTS

<u>Method of Scoring</u>	<u>N</u>	<u>No. of Items</u>	<u>Reliability</u>
Unclustered	77	100	.62
Unclustered, selected items	40	59	.75
Profile scoring, selected items	40	59	.67

Externals.

Method. Substantially the same procedures were used here as in the AS section. Self descriptions of the five key persons, were compared with the self descriptions of 40 others by the unclustered method. Self descriptions of two key persons, chosen at random from the original five, were compared with 40 others, using the profile-scoring method. Only two of the original five key persons were used in this part of the study, because of the computational time and labor involved.

Results. The results are summarized in Table 6, which includes RS reliabilities from the previous section for purposes of comparison. With selection of items, all RS unclustered scores attained reliability greater than zero, but not much greater. Cluster scoring did not appear to have a beneficial effect.

Table 6
RELIABILITIES OF RS OF THE ORIGINAL, SHORTENED, AND
PROFILE SCORED TESTS FOR FIVE KEY PERSONS

<u>Key Person</u>	<u>100-Item Unclustered (N = 77)</u>	<u>59-Item Unclustered (N = 40)</u>	<u>59-Item Profile Scored (N = 40)</u>
1	-.226	.109	-.624
2	.068	.337	---
3	.264	.027	---
4	-.571	.099	---
5	-.302	.370	.047

Conclusions. For these items, RS shows extremely low reliability. Selecting items which correlate raised reliability, but not enough to make the RS score on this instrument useful. The slight improvement in reliability may be

due to retaining more reliable items, but is probably also due to increasing the homogeneity of the items. Shortening of the test, of course, tends to affect reliability adversely.

Profile scoring does not seem to help. This conclusion, for the five-item forced-choice block, is supported by preliminary indications from a mathematical study by Warrington.

Reliability over Perceives

As stated in the introduction, the second method of testing reliability called for the demonstration of consistency over perceives. An ancillary study was conducted by Fiedler in which this hypothesis along with several others was tested.

Captains of seventeen university intramural basketball teams were asked to complete test blanks for (1) self description, (2) prediction of best cooperator, and (3) prediction of second best cooperator. The Assumed Similarity to the best cooperator was then correlated with the assumed similarity to the second best cooperator over the seventeen players. Rho was .58. This compares favorably with the split-half reliability of ASp on the original 100-item instrument ($r = .618$).

While the small N must be kept in mind, it is tentatively concluded--pending study with a larger N--that a second phase of the reliability of ASp, reliability over perceives, has been demonstrated. The result does tend to show that the person who assumes similarity in one case also does so when perceiving another, but the generality of this attitude is a problem for further study. The fact that ASp does not depend on a specific perceiver makes it more significant as an expression of some personality characteristic of the respondent.

While correlation of the two ASp measures was high, the correlation must be influenced by the degree of preference the perceiver has for the two perceives. Thus, the correlation between ASp for best cooperator and ASp for, say, fifth cooperator would presumably be less than the correlation of .58 between ASp for best cooperator and second best cooperator. Further, it seems likely that the distance between the best cooperator and the second best cooperator as used in this study is not the same for every subject.

Summary

A study was performed to investigate the reliability of certain interpersonal perception measures, using data gathered in a study on basketball teams (3).

The two classes of variables were (1) Assumed Similarity, defined as the correlation between a person's self description and his prediction of the responses of another on a set of items, or the correlation between two predictions of others that one person has made, and (2) Real Similarity, defined as the correlation of different persons' descriptions or predictions on a set of findings. The major findings were:

1. Measures of Assumed Similarity derived from the basketball instrument possess reliability sufficiently high to detect group differences and allow for exploratory studies. The AS tendency is reliable over sets of items and over perceivers.
2. The reliability of Assumed Similarity measures can be materially improved by selecting items that are unambiguous and correlate with other items. It was not improved by profile scoring.
3. Measures of Real Similarity on the present instrument were very unreliable and cannot be expected to correlate with criteria. The reliability of these measures increased slightly but not sufficiently by item selection. Profile scoring was of no help. However, it should be noted that the profile scoring procedure was not finally tested here, having been limited to five-item blocks and clusters of low internal consistency.

The study implies need for better measures of real similarity. Assumed similarity is established as a trait which can be measured and whose psychological properties should be further investigated. Pretesting of items is advised for further instruments.

References

1. Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.
2. Cronbach, L. J., and Gleser, Goldine C. Similarity between persons and related problems of profile analysis. Champaign-Urbana, Illinois: Univ. of Illinois, 1952. (Mimeographed, Technical Report No. 2, Contract N6ori-07135 between the University of Illinois and the Office of Naval Research.)
3. Fiedler, F. E., Hartmann, W., and Rudin, S. A. The relationship of interpersonal perception to effectiveness in basketball teams. Champaign-Urbana, Illinois: Univ. of Illinois, 1952. (Mimeographed, Technical Report No. 3, Contract N6ori-07135 between the University of Illinois and the Office of Naval Research.)
4. Guttman, L. A basis for analyzing test-retest reliability. Psychometrika, 1945, 10, 255-282.
5. Stephenson, W. A. Statistical approach to typology: the study of trait universes. J. clin. Psychol., 1950, 6, 26-38.



UNIVERSITY OF ILLINOIS-URBANA



3 0112 084224671